Design of multimodal antibiotics against intracellular infections using deep learning

Angela Cesaro[1-4], Fangping Wan[1-4], Marcelo D. T. Torres[1-4], and Cesar de la Fuente-Nunez[1-4]*

Affiliations

[1]Machine Biology Group, Departments of Psychiatry and Microbiology, Institute for Biomedical Informatics, Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America.

[2]Departments of Bioengineering and Chemical and Biomolecular Engineering, School of Engineering and Applied Science, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America.

[3]Department of Chemistry, School of Arts and Sciences, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America.

[4]Penn Institute for Computational Science, University of Pennsylvania; Philadelphia, Pennsylvania, United States of America.

*Corresponding author. Email: cfuente@upenn.edu

Abstract

The rise of antimicrobial resistance has rendered many treatments ineffective, posing serious public health challenges. Intracellular infections are particularly difficult to treat since conventional antibiotics fail to neutralize pathogens hidden within human cells. However, designing molecules that penetrate human cells while retaining antimicrobial activity has historically been a major challenge. Here, we introduce APEX$_{DUO}$, a multimodal artificial intelligence (AI) model for generating peptides with both cell-penetrating and antimicrobial properties. From a library of 50 million AI-generated compounds, we selected and characterized several candidates. Our lead, Turingcin, penetrated mammalian cells and eradicated intracellular *Staphylococcus aureus*. In mouse models of skin abscess and peritonitis, Turingcin reduced bacterial loads by up to two orders of magnitude. In sum, APEX$_{DUO}$ generated multimodal antibiotics, opening new avenues for molecular design.

Introduction

Antimicrobial resistance is a critical global health threat, with nearly three million infections caused by resistant pathogens occurring annually in the United States alone, resulting in over 35,000 deaths and treatment costs exceeding $4.6 billion [1,2]. As pathogenic bacteria evolve diverse resistance mechanisms, they can also employ strategies to evade both host immune defenses and conventional antibiotics. Among these strategies is intracellular pathogenicity, wherein certain bacterial species, such as *Staphylococcus aureus*, invade mammalian cells, replicate intracellularly, and subsequently re-emerge to infect new tissues. This process can lead to severe conditions, including sepsis [3–5]. Addressing this challenge requires novel therapeutics that combine multiple and complementary biological activities.

Traditional drug discovery strategies have proven time-consuming, costly, and increasingly inadequate for addressing the escalating resistance problem. Moreover, these methods predominantly target extracellular bacteria, failing to address the pressing need for therapies capable of eradicating intracellular pathogens [6,7].

Here, we introduce APEX$_{\text{DUO}}$, a deep learning model designed to generate peptide candidates with both cell-penetrating and antimicrobial properties. Cell penetrating peptides (CPPs) facilitate the translocation of biomolecules, including therapeutics, into mammalian cells, while antimicrobial peptides (AMPs) offer broad-spectrum antimicrobial activity against diverse pathogens and may modulate immune responses [8–10]. Our objective was to create novel, multimodal molecules capable of penetrating host cells and eradicating intracellular bacteria, thereby addressing critical gaps in the treatment of persistent intracellular infections.

Using APEX$_{\text{DUO}}$, we generated a library of 50 million peptide candidates, synthesized and experimentally tested 60 sequences —collectively termed Turingcins in honor of Alan Turing. Among these, one peptide, Turingcin, exhibited remarkable efficacy in eliminating intracellular *S. aureus* infections *in vitro* and in preclinical mouse models.

Results

Overview of APEX$_{\text{DUO}}$

To design peptides with dual cell-penetrating and antimicrobial properties, we leveraged a deep generative model, specifically a maximum mean discrepancy variational autoencoder (MMD-VAE), to produce novel peptide sequences. We then applied deep learning-based predictors to assess both cell-penetrating and antimicrobial activities, identifying candidates with the desired multimodal profiles (**Fig. 1A**). From the 50 million peptides generated, we applied stringent filtering criteria to select 60 top-ranked compounds for synthesis and characterization (see "**Peptide selection**" subsection of the **Methods**). These criteria included excluding peptides with high median MIC values, low cell-penetration probabilities, duplicates, and peptides already present in in-house datasets and public databases, as well as minimizing sequence redundancy. After applying these filters, 52 soluble peptides—collectively termed Turingcins—were selected for experimental testing **(Table S1)**.

Composition and physicochemical features of multimodal candidates

To investigate the distribution and similarity of the generated multimodal candidates, we used first token embeddings from the pretrained protein large language model ESM2 to represent peptides from various sources, including non-AMPs/CPPs, AMPs/CPPs, generated peptides, and synthesized peptides. These representations were plotted in 2D using Uniform Manifold Approximation and Projection (UMAP), revealing that the generated multimodal candidates overlapped with known AMPs/CPPs. This overlap suggests their potential to exhibit both antimicrobial and cell-penetrating properties **(Fig. 1B)**.

An analysis of amino acid composition revealed distinctive features of Turingcins. These peptides showed slight enrichments in isoleucine, lysine, leucine, and asparagine frequencies, contributing to their higher net positive charge relative to common AMPs and CPPs, which generally exhibit lower lysine levels. Conversely, Turingcins displayed reduced levels of cysteine, aspartic acid, glutamic acid, and glycine, as well as lower arginine content compared to CPPs, and lower valine content compared to AMPs. Notably, the synthesized and validated Turingcins lacked cysteine, aspartic acid, and proline residues altogether **(Fig. 1C)**.

Regarding physicochemical properties, the multimodal candidates demonstrated lower normalized hydrophobicity values compared to known AMPs and CPPs but exhibited intermediate amphiphilicity index values—an important trait for peptides designed to penetrate biological membranes **(Figs. 1D- G)**. This suggests their membrane interactions are less intense than those of highly amphiphilic peptides. The Turingcins showed intermediate values for amphiphilicity, hydrophobicity, and disordered conformation propensity, bridging the characteristics of AMPs and CPPs. However, their net charge, hydrophobic residue angle, and isoelectric point aligned more closely with CPPs, while other features more closely resembled AMPs **(Figs. 1D- g; fig. S1).**

These distinct physicochemical and compositional features set Turingcins apart from previously described peptide classes, highlighting their unique potential as multimodal therapeutics.

In vitro **antimicrobial activity of multimodal candidates**

The antimicrobial efficacy of the designed Turingcins was initially evaluated against twelve clinically relevant bacterial strains, encompassing both Gram-positive and Gram-negative species, with a focus on members of the ESKAPEE group—recognized by the World Health Organization as major public health threats [11] **(Fig. 2B)**. A remarkable 90.38% (47/52) of Turingcins exhibited activity against at least one pathogen with MIC_{100} values (minimal inhibitory concentrations resulting in complete inhibition of microbial growth) ranging from 0.78 to 50 μmol L^{-1}. Among these, 15 peptides demonstrated activity against *S. aureus* ATCC 12600.

Due to its ability to invade host cells, establish persistent colonization, and cause recurrent infections by replicating within both phagocytic and non-phagocytic cells [12], *S. aureus* was selected as the standard intracellular pathogen for experimental characterization in this study.

Mechanism of action of Turingcins

To explore the antimicrobial mechanisms of the 15 Turingcins active against *S. aureus* ATCC 12600, we assessed their ability to depolarize bacterial membranes, a typical target of AMPs through non-specific lipid bilayer interactions [13] **(Fig. 2C; fig. S2A)**. Using the voltage-sensitive dye $DiSC_3$-5, which fluoresces upon membrane depolarization, we identified significant perturbations in membrane potential in the case of five peptides (Turingcins 34, 35, 37, 57, and 58), representing 33% of those tested.

In contrast, nine active peptides (53%; Turingcins 2, 8, 12, 15, 17, 22, 23, 46, and 47) exhibited minimal or no significant fluorescence changes, suggesting alternative antimicrobial mechanisms. Interestingly, Turingcin-28 reduced the fluorescence of the probe, a phenomenon potentially linked to its high tryptophan content (6 out of 18 residues, 33%), the highest among the tested sequences **(Fig. 2C, fig. S2A)**.

This diversity in mechanisms of action highlights the algorithm's ability to generate peptides with multifaceted antimicrobial strategies, reflecting a versatile approach to addressing infections.

In vitro cytotoxicity and hemolytic activity

Evaluating the potential toxicities of Turingcins is a critical step in assessing their bioactivity and therapeutic potential. To assess their toxicity, we conducted dose-response 3-(4, 5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide (MTT) cytotoxicity assays using immortalized human keratinocytes (HaCaT) and phorbol 12-myristate 13-acetate (PMA)-differentiated human monocytes (THP-1), modeling macrophages [14,15]. These assays assessed the biocompatibility of 43 antimicrobial Turingcins with both non-phagocytic and phagocytic cells **(Fig. 2D; figs. S2B and S2C)**. CC_{50} values (the concentration causing 50% cell death) were determined through non-linear regression of dose-response data (3.12–50 μmol L^{-1}) and compared to their corresponding MIC_{100} values **(Fig. 2D; figs. S2B and S2C)**.

In HaCaT cells, 9.3% of Turingcins (4 out of 43) showed CC_{50} values below 12.5 μmol L^{-1}, compared to only 3% (1 out of 43) in PMA-differentiated THP-1 cells, indicating greater susceptibility of keratinocytes than macrophages to Turingcins. Nearly all peptides exhibited CC_{50} values at least twice their MIC_{100}, except Turingcin-32, whose antimicrobial effective dose (50 μmol L^{-1}) exceeded its CC_{50} against HaCaT cells (34.84 μmol L^{-1}).

To further assess safety, we evaluated the hemolytic activity of Turingcins by exposing red blood cells (RBCs) to the peptides, a standard method for toxicity evaluation of antimicrobial agents [15]. Four of the 43 Turingcins tested (9.3%) displayed hemolytic concentration causing 50% cell death (HC_{50}) at values below 12.5 μmol L^{-1} **(Fig. 2D; fig. S2D)**.

These findings highlight the promising safety profile of $APEX_{DUO}$-designed compounds. Most Turingcins demonstrated CC_{50} or HC_{50} values that exceeded their MIC_{100} by 2 to over 50 times and surpassed 100 times in the case of RBCs. These findings underscore the success of $APEX_{DUO}$ in balancing antimicrobial potency with minimal host cell toxicity.

Secondary structure of Turingcins

The secondary structures of short peptides are inherently dynamic, often shifting between disordered and ordered states at hydrophobic-hydrophilic interfaces. For AMPs, this structural plasticity is crucial for their interaction with biological membranes, as it underlies their functional mechanisms of action [16]. Likewise, recent studies indicate that the secondary structural properties and conformations of CPPs profoundly influence their cellular uptake pathways and overall internalization efficiency, underscoring the importance of peptide structure in determining both antimicrobial and cell-penetrative activities [17].

To characterize the secondary structure of the synthesized Turingcins, Circular Dichroism (CD) spectroscopy was performed in various environments: water, trifluoroethanol (TFE) in water mixture (3:2, v:v), methanol (MeOH) in water mixture (1:1, v:v), and sodium dodecyl sulfate (SDS) in water (10 mmol L^{-1}). The TFE/water and MeOH/water mixtures were used to induce distinct secondary structures: TFE usually promotes α-helical conformations by dehydrating peptide amide groups and facilitating intramolecular hydrogen bonds, while MeOH favors β-sheet conformations by interacting with the amide groups of the backbone and favoring intermolecular side chains interactions. SDS micelles were included to mimic the bilayer environment of biological membranes and provide insight into how the peptides might behave *in vivo* [18–20].

All Turingcins were analyzed using CD spectroscopy over a wavelength range of 260 to 190 nm **(Fig. 2E; fig. S3).** Secondary structure fractions were estimated using the BeStSel server [21]. In water, most peptides were predominantly unstructured. Notably, however, the inactive peptides Turingcins 1, 7, and 21 showed a pronounced tendency to adopt α-helical structures (31.4%, 39%, and 29.5%, respectively). In TFE and water mixture, unstructured conformations remained dominant, whereas in MeOH and water mixture, peptides displayed enhanced β-like structural features. In the presence of SDS micelles in water, which are intended to simulate membrane-like environments, the unstructured state generally predominated, with certain peptides exceeding 65% unstructured content **(Fig. 2E; fig. S3).** In summary, Turingcins are generally unstructured and tend to β-like conformations in β-inducer medium.

Intracellular antimicrobial activity of Turingcin

To evaluate the intracellular antimicrobial activity of the most promising Turingcins, we developed a model using HaCaT and PMA-differentiated THP-1 cells infected with intracellular *S. aureus* ATCC 12600. From the panel of Turingcins, we selected nine highly active candidates against *S. aureus*, testing them at three concentrations (12.5, 25, and 50 μmol L^{-1}), including doses known to elicit mild cytotoxic effects **(Figs. S4A and S4B)**. This strategy allowed us to identify time-dependent trends while excluding conditions where toxicity obscured antimicrobial efficacy.

Most of the tested Turingcins effectively reduced the intracellular bacterial load, with particularly strong effects observed at earlier time points (1 and 3 hours), likely due to peptide degradation, and at higher concentrations (25 and 50 μmol L^{-1}) **(Figs. S4A and**

S4B). To ensure reliable results while accounting for potential cytotoxicity, we prioritized the most promising candidate, Turingcin-46. This peptide demonstrated robust antimicrobial activity against *S. aureus* while maintaining favorable cytotoxicity profiles in HaCaT cells and PMA-differentiated THP-1 cells. At non-cytotoxic concentrations, Turingcin reduced the intracellular bacterial burden by 1 and 3 orders of magnitude in HaCaT and PMA-differentiated THP-1 cells, respectively, compared to untreated controls **(Figs. 3A and B)**. In addition to the results observed with Turingcin-46 (hereafter referred to as Turingcin), Turingcin-47 also demonstrated the ability to reduce the intracellular bacterial load, although its effects were significant only at 1 hour and were less pronounced.

To further examine Turingcin's cell-penetration capabilities and intracellular localization, we conjugated the peptide with Cyanine-5 (Cy5) at the N-terminus and applied it to both infected and uninfected cells. Two concentrations of Cy5-Turingcin-46 (25 and 50 µmol L^{-1}) were tested at 1- and 3-hours post-treatment. Afterward, cells were fixed and stained with Hoechst (blue) to label nuclei and WGA (green) to visualize cell walls, then analyzed by fluorescence microscopy. Substantial intracellular accumulation of Cy5-Turingcin (purple signal) was observed in both infected **(Figs. S4C and D)** and uninfected cell monolayers **(Figs. 3C and D)**, compared to untreated controls.

These findings confirm Turingcin's ability to penetrate both non-phagocytic and phagocytic eukaryotic cells and effectively eliminate intracellular *S. aureus*.

Ala-scan screening of Turingcin to elucidate structure-function relationships

To dissect the contributions of individual amino acids to Turingcin's biological activity, we performed alanine-scan screening. Each residue in Turingcin was systematically replaced with alanine, generating 18 single-substitution variants **(Fig. 4; Table S2, figs. S5 and S6)**. Alanine, with its short aliphatic side chain, was chosen to preserve peptide backbone length while minimizing disruptive interactions between side chains [22,23].

Analysis of the antimicrobial and cytotoxic properties of these alanine-substituted variants revealed that mutations at positions 1 (F), 3 (W), 4 (W), 5 (L), 10 (L), 19 (R), 20 (R), and 21 (K) significantly increased the MIC_{100} against *S. aureus* ATCC 12600, rendering the peptides inactive **(Fig. 4A; fig. S5A)**. Substitutions at positions 15 (I) and 17 (I) also diminished antimicrobial potency, albeit to a lesser degree. These findings highlight the importance of hydrophobic and aliphatic residues such as tryptophan, phenylalanine, leucine, and isoleucine, as well as positively charged arginine and lysine, in mediating antimicrobial effects. Notably, the changes in cationic residues close to the C-terminus were more impactful than close to the N-terminus where cationic residues are surrounded by hydrophobic and aliphatic residues.

Some of the residues affecting MIC_{100} values also impacted cytotoxicity **(Fig. 4A; figs. S5B and S5C)**. Mutations leading to a loss of antimicrobial activity generally reduced cytotoxicity in PMA-differentiated THP-1 cells (e.g., A4W, A5L, A20R, and A21K). All the derivatives had similar activity profiles or were more toxic than Turingcin against keratinocytes. The wild-type peptide did not present high hemolytic activity, and the Ala-scan screening showed that substitutions that led to enhanced mean hydrophobicity values correlated with a lower propensity for hemolysis **(Fig. 4A; table S3 and fig. S5D)**.

Circular dichroism analysis showed that Turingcin and its alanine-substituted variants generally adopted similar secondary structures **(Fig. 4B; fig. S6)**. In aqueous solution, the peptides were predominantly unstructured, with some β-like conformations. In TFE and water mixture, variants that lost antimicrobial activity showed an increased tendency to form α-helical structures (up to 37%). Nevertheless, in MeOH and water mixture and SDS micelles, both wild-type and mutated sequences displayed similar proportions of unstructured and β-like conformations **(Fig. 4B; fig. S6)**.

To examine how these structural changes affected interactions with biological membranes, we employed two fluorescent probes—$DiSC_3$-5 and $DiBAC_4$-3 [24,25]—to evaluate membrane depolarization in bacterial and eukaryotic cells, respectively. Turingcin had a modest ability to depolarize the bacterial membrane, a trait shared by most alanine-substituted variants **(Fig. 4C; fig. S5E)**. Interestingly, A15I- and A17I-substituted Turingcin displayed enhanced depolarizing activity against *S. aureus* membranes **(Fig. 4C; fig. S5E)**.

The peptides exhibited distinct effects on eukaryotic cell membranes. In HaCaT keratinocytes, Turingcin did not alter membrane potential, whereas most alanine-substituted derivatives induced hyperpolarization (indicated by decreased fluorescence), particularly those mutated at the N- and C-termini. Conversely, A7K- and A15I-substituted variants increased fluorescence, indicating membrane depolarization. In PMA-differentiated THP-1 macrophages, wild-type Turingcin induced membrane depolarization, whereas most alanine-substituted variants showed reduced depolarizing effects **(Fig. 4D; figs. S5F-G)**. One exception was A7K-Turingcin, which caused significantly greater depolarization than the wild-type sequence. These results underscore the cell-type specificity of peptide-membrane interactions influenced by phospholipid composition.

Despite retaining some antimicrobial properties and cytotoxic profiles, alanine-substituted variant A7K-Turingcin failed to reduce the intracellular bacterial burden in HaCaT and THP-1 cells when tested at non-cytoxic concentrations **(Fig. 4E).** Conversely, A16-Turingcin preserved intracellular activity in macrophages, although its effectiveness was diminished compared to the lead wild-type sequence **(Fig. 4E)**. This suggests that these variants lack key sequence features necessary for the full multimodal activity of the wild-type peptide.

In summary, alanine-scan screening revealed the importance of Turingcin's unique combination and distribution of amino acid residues in mediating its membrane interactions and multifunctional biological activity. This analysis highlights the nuanced interplay between sequence composition and peptide function in computationally designed molecules.

Resistance to proteolytic degradation

Before conducting *in vivo* mouse experiments, we assessed Turingcin's stability in the presence of serum proteases. After one hour of incubation, approximately 50% of the peptide remained intact. We attribute this stability to the presence of hindered hydrophobic residues such as phenylalanine and tryptophan that may further reduce protease accessibility **(Fig. S7A).**

Anti-infective efficacy of Turingcin in animal models

To investigate Turingcin's antimicrobial efficacy in complex biological environments, we tested it in two mouse models: a skin abscess model and a peritonitis/sepsis model **(Fig. 5; figs. S7B-D)**.

In the skin abscess model, a 20 μL suspension containing 10^7 *S. aureus* ATCC 12600 cells in phosphate-buffered saline (PBS) was applied to a wound on the back of the mice [11,26–28] **(Figs. 5A and B; fig. S7B)**. A single dose of Turingcin was administered directly to the infected area. After 24 hours, the bacterial count decreased by more than two orders of magnitude compared to the untreated control group, with Turingcin demonstrating slightly greater potency than gentamicin **(Fig. 5B)**. Confocal microscopy of skin samples treated with Cy5-labeled Turingcin confirmed the peptide's cellular penetration capability, showing fluorescence within the stratified squamous epithelium **(Fig. 5C; fig. S7C)**. Throughout the experiments, no significant changes in body weight, a general indicator of potential toxicity, were detected **(Fig. S7B)**.

Next, we evaluated Turingcin in a peritonitis/sepsis mouse model **(Figs. 5D and E; fig. S7D)**, which allows for the evaluation of intracellular antimicrobial activity [29,30]. Mice were intraperitoneally infected with 100 μL containing 5×10^{10} *S. aureus* ATCC 12600 cells *per* mL. Two hours post-infection, three doses of Turingcin were administered intraperitoneally at 1.5-hour intervals to mitigate peptide degradation (as confirmed by prior proteolytic stability studies).

One day after treatment, Turingcin significantly reduced bacterial burdens in multiple organs, including the liver, kidneys, spleen, lungs, and heart, with reductions ranging from 1.2 to 1.8 orders of magnitude. Notably, Turingcin outperformed gentamicin (positive control) in the liver, spleen, and lungs, suggesting possible organ-specific peptide accumulation. Analysis of peritoneal fluid showed a substantial decrease in *S. aureus* counts, including both extracellular and intracellular bacteria, exceeding one order of magnitude. Importantly, no significant changes in body weight were observed, further supporting the peptide's safety at the tested dose **(Fig. S7D)**.

Overall, these *in vivo* findings highlight Turingcin's potent antibiotic activity under physiological conditions and highlight its potential as a multimodal therapeutic agent against intracellular infections.

Discussion

Eliminating intracellular pathogens remains a significant therapeutic challenge. Traditional antibiotics often fail to reach sufficient concentrations within host cells to effectively eradicate intracellular bacteria, and achieving therapeutic intracellular levels frequently requires high, potentially toxic doses [31].

While multimodal models are now widely applied in domains such as chatbots, their use in molecular design has been limited due to the complexity of integrating multiple functionalities into a single molecule. Multimodal molecules offer a promising strategy

to overcome these barriers, but their design and optimization using conventional methods have proven difficult.

$APEX_{DUO}$ addresses this challenge by integrating AI models with predictive tools for antimicrobial and cell-penetrating activities. This approach yielded Turingcins, a class of novel peptides with unique compositions straddling the known CPP and AMP chemical spaces. From the millions of molecules generated by $APEX_{DUO}$, we prioritized sequences with antimicrobial activity against *S. aureus*, a key intracellular pathogen. Among these, Turingcin emerged as the most promising candidate, significantly reducing intracellular *S. aureus* loads in keratinocytes and macrophages.

Mutational analysis identified sequence elements critical for Turingcin's antimicrobial efficacy and cytotoxic selectivity, correlating these features with the peptide's ability to penetrate mammalian cells, eliminate intracellular pathogens, and maintain a favorable cytotoxic profile. Furthermore, *in vivo* experiments demonstrated Turingcin's robust antimicrobial efficacy in two preclinical mouse models, underscoring its therapeutic potential. In summary, $APEX_{DUO}$ demonstrates that AI can be used to design multimodal peptides, opening new avenues for tackling the formidable challenge of intracellular infections.

Limitations of the study

Despite its success, $APEX_{DUO}$ has limitations. The antimicrobial predictor did not incorporate datasets specifically annotated for intracellular activity, which may have constrained the accuracy of predictions for this context. Similarly, limited publicly available data on CPPs restricted the predictor's training. Future efforts to curate large datasets with intracellularly active peptide and cell-penetration annotations will enhance predictive models and accelerate the discovery of improved intracellular antibiotics. Additionally, our study focused on peptides with confirmed antimicrobial activity for downstream intracellular assays, excluding purely cell-penetrating candidates. Broader testing could provide further insights into the full spectrum of $APEX_{DUO}$-generated molecules and identify additional therapeutic candidates.

Materials and Methods

Datasets

Peptides used for training the deep generative model were curated from UniRef50 [32]. We filtered out protein sequences having more than 50 amino acids. The peptides we downloaded from UniRef50 were divided into a training set and a validation set with 1,418,011 and 354,503 peptides, respectively. AMPs used for computational analysis were retreived from public databases, including APD3 [33], DBAASP [34] and DRAMP [35], as well as our in-house peptides having minimum inhibitory concentration ≤128 μmol L^{-1} against at least one bacterial strain [36]. CPPs used for training and computational analysis were retreived from CPPsite 2.0 [37,38], SkipCPP-Pred [39] and MLCPP [40]. We also curated toxic peptides from SATPdb [41] and labeled them as negative data so that the CPP predictor can balance the cell-penetrating property and toxicity and help to retrieve peptides with good cell penetration and relatively low toxicity. After cleaning the non-canonical

peptides and removing duplicates, we obtained 1,217 CPPs and 5,293 non-CPPs or toxic peptides for CPP predictor building.

Variational autoencoder

We use a maximum mean discrepancy variational autoencoder (MMD-VAE) as our peptide sequence generator [42]. The MMD-VAE consists of an encoder $E(\cdot)$ and a decoder $D(\cdot)$, both typically implemented as neural networks. The encoder and decoder form a compression and de-compression process that enables representation learning, distribution learning and novel data generation. In our scenario, the inputs to the encoder are peptide sequences. For a peptide sequence with l amino acids, two special characters indicating the start and end of the sequence are added at the beginning and end, respectively. One-hot encoding is then used to numerically represent this sequence. Let x denotes this representation, the encoder takes x and generates a continuous vector (i.e., a latent code) $v = E(x) \in \mathbb{R}^d$, where d stands for the dimension of v. The decoder uses the latent code v to generate a peptide sample $x' = D(v)$ that resembles x. Given a peptide set to train the MMD-VAE and let $p_{data}(x)$ denote the empirical distribution of training peptides, the minimization goal of MMD-VAE can be written as

$$loss = -\mathbb{E}_{p_{data}(x)}\left[\mathbb{E}_{p(v|x)}[\log p(x|v)]\right] + \mathrm{MMD}(p(v|x), p_{prior}(v)) \qquad \text{Eq.1}$$

where $\mathbb{E}[\cdot]$ stands expectation, $p(v|x)$ is the posterior distribution of the latent codes given xs from $p_{data}(x)$, MMD is the maximum mean discrepancy that measures the distance of two probability distributions, and $p_{prior}(v)$ is the prior distribution of v. In our case, the log-likelihood $\log p(x|v)$ in Eq.1 is equal to the negative cross entropy between x from $p_{data}(x)$ and its reconstructed version by $D\big(E(x)\big)$ the MMD-VAE. In this sense, Eq.1 can be re-written as

$$loss = \mathbb{E}_{p_{data}(x)}\left[\mathrm{CE}(x, D(E(x))\right] + \mathrm{MMD}(p(v|x), p_{prior}(v)),$$

where $\mathrm{CE}(\cdot,\cdot)$ is the cross entropy. Optimizing the parameters of $E(\cdot)$ and $D(\cdot)$ to minimize $\mathbb{E}_{p_{data}(x)}\left[\mathrm{CE}(x, D(E(x))\right]$ ensures the MMD-VAE to capture the distribution of training peptide sequences (hence, the decoder can generate peptides resemble the training ones) while the optimization of $E(\cdot)$ and $D(\cdot)$ for $\mathrm{MMD}(p(v|x), p_{prior}(v))$ makes the latent codes vs close to the prior distribution (hence, making it easier for the users to sample latent codes from the prior distribution and generate new peptide samples). In our practice, the prior distribution is the zero-mean, unit-variance multivariate Gaussian distribution. The encoder consists of a recurrent neural network, attention neural network, and a fully-connected neural network layer for extracting residue-wise features, creating a weighted summation of residue-wise features (i.e., a peptide level feature), and transforming the peptide level feature to the latent code, respectively. The decoder is a recurrent neural network with the same architecture as that from the encoder followed by a fully-connected neural network layer to transform the latent code to the residue type probabilities per amino acid position. We used a learning rate of 0.0001 to train the MMD-VAE 1,000 epochs on the training set. The number of recurrent neural network layers, the number of hidden units in the recurrent neural network, the dimension of the latent code, batch size, weight decay strength, and the best epoch were chosen by the $loss$ of the trained model evaluated on the validation set.

Peptide generation

Instead of directly sampling latent codes to generate peptide samples, we followed the previous practice to fit a kernel density estimator (KDE) to the functional peptides. That is, we obtained the latent codes of our in-house peptides and publicly available CPPs. A KDE (implemented by scikit-learn with default hyperparameter) was used to fit the distribution of these latent codes. We then sampled the latent codes from the KDE and decoded the sampled codes by the decoder to obtain the generated peptide sequences. We sampled 50M latent codes and generated the corresponding sequences for further filtering by the antimicrobial and cell-penetrating peptide predictors.

Antimicrobial peptide predictor

The antimicrobial peptide we used is APEX, a multi-bacterial multi-strain antimicrobial peptide predictor previously been built for mining antibiotic encrypted peptides from proteomes of extinct organism. For each generated peptide, we used APEX to obtain the median value of predicted minimum inhibitory concentrations (MICs) against multiple bacterial strains.

Cell-penetrating peptide predictor

We used deep learning to build the binary classifier to predict the cell-penetrating property of peptide sequences. Our predictor has similar architecture to that of the encoder of MMD-VAE except that we added one more fully-connected neural network layer to the hidden peptide features to the cell-penetrating probabilities. We set the learning rate = 0. 0001, batch size = 128, the number of training epochs = 100, and trained each model five times under different seeds to create an ensemble version of it (i.e., we averaged the prediction result from each model). Hyperparameters (i.e., the number of recurrent neural network layers, the hidden unit size, and weight decay strength) of the CPP predictor were selected based on the mean of the area under the receiver operating characteristic under the five-fold cross validation. The CPP predictors from the top five hyperparameter combinations were selected (i.e., 25 models = 5 hyperparameter combinations × 5 copies under different random seeds) for the CPP peptide screening.

Peptide selection

We used the following rules to filter and selection the desired peptides from 50M generated peptides from the MMD-VAE.

(1) Generated peptides having median MIC >32 μmol L^{-1} were filtered out from our selection.
(2) After screening the CPP probabilities for the 50M generated peptides, we calculated the mean and standard deviation of the predicted probabilities among the 50M sequences. Generated peptides passing 3-sigma rule (i.e., CPP prediction ≥ mean of CPP prediction + 3× standard deviation of CPP prediction) were kept.

(3) Generated peptides appeared in our in-house peptide data or training CPPs were removed.

(4) Remaining generated peptides were ranked by median MIC, and we used Striped Smith–Waterman algorithm to perform pair-wise sequence alignment and calculate the sequence similarity between two peptides. We kept the peptide with lower median MIC when a pair of peptides having a sequence similarity > 0.5.

Physicochemical properties extraction

The twelve physicochemical properties of the peptides, including normalized hydrophobic moment, normalized hydrophobicity, net charge, amphiphilicity index, isoelectric point, penetration depth, tilt angle, disordered conformation propensity, linear moment, tendency for in vitro aggregation, angle between hydrophobic residues, and propensity for PPII coil, were obtained using the DBAASP server [43]. The hydrophobicity scale used was the Eisenberg and Weiss scale.

Peptide synthesis

The peptides used in the experiments were procured from AAPPTec and produced through solid-phase synthesis employing the Fmoc strategy.

Bacterial strains and growth conditions

In this study, we worked with the following pathogenic bacterial strains: *Staphylococcus aureus* ATCC 12600, *Staphylococcus aureus* ATCC BAA-1556 (methicillin-resistant strain), *Enterococcus faecalis* ATCC 700802 (vancomycin-resistant strain), *Enterococcus faecium* ATCC 700221 (vancomycin-resistant strain), *Listeria monocytogenens* ATCC 19111, *Salmonella enterica* ATCC 9150, *Escherichia coli* ATCC 11775, *Escherichia coli* AIC221 [*Escherichia coli* MG1655 phnE_2::FRT (control strain for AIC222)], *Escherichia coli* AIC222 [*Escherichia coli* MG1655 pmrA53 phnE_2::FRT (polymyxin resistant; colistin-resistant strain)], *Acinetobacter baumannii* ATCC 19606, *Klebsiella pneumoniae* ATCC 13883, and *Pseudomonas aeruginosa* PAO1. All strains were cultured in Luria-Bertani (LB) broth or on LB agar, except for *L. monocytogenes*, which was cultured in Brain Heart Infusion (BHI) medium, and *S. aureus*, which was grown on mannitol salt agar plates. Bacterial cultures for all experiments were started from a single isolated colony and grown overnight (16 hours) in liquid medium at 37 °C. The next day, the cultures were diluted 1:100 into fresh medium and incubated at 37 °C until reaching the mid-logarithmic growth phase.

Minimal inhibitory concentration assays

Broth microdilution assays were conducted to determine the MIC_{100}, the minimum inhibitory concentration required to achieve complete bacterial elimination, for each peptide. Peptides were prepared in sterile water and serially diluted twofold in untreated polystyrene 96-well plates, with final concentrations ranging from 0 to 100 μmol L^{-1}.

Bacterial cultures, adjusted to 4×10^6 CFU mL^{-1} in LB or BHI medium, were combined with the peptide solutions in equal volumes. After 24 hours of incubation at 37 °C, the MIC_{100} was identified as the lowest peptide concentration that prevented visible bacterial growth.

Bacterial membrane depolarization assays using $DiSC_3$-5

A cytoplasmic membrane depolarization assay was carried out using the membrane potential-sensitive dye 3,3'-dipropylthiadicarbocyanine iodide ($DiSC_3$-5, Sigma 43608). Mid-logarithmic phase *S. aureus* ATCC 12600 was washed and adjusted to an optical density of 0.05 at 600 nm in HEPES buffer (pH 7.2) supplemented with 20 mmol L^{-1} glucose and 0.1 mol L^{-1} KCl. The bacterial suspension (100 µL per well) was incubated with 10 µmol L^{-1} $DiSC_3$-5 for 15 minutes. This allowed fluorescence to stabilize as the dye incorporated into the bacterial membrane. The peptides were subsequently added at concentrations corresponding to their MIC_{100} values in a 1:1 ratio with the bacterial suspension. Fluorescence changes (λ_{ex} = 622 nm, λ_{em} = 670 nm) were recorded over 45 minutes to monitor membrane depolarization.

Eukaryotic cell culture and cytotoxicity assays

Immortalized human keratinocytes (HaCaT, sourced from Deutsches Krebsforschungszentrum - DKFZ [44]) and human monocytes (THP-1, ATCC TIB-202) were maintained in high-glucose Dulbecco's modified Eagle's medium (DMEM) and Roswell Park Memorial Institute 1640 (RPMI), respectively. Both media were enriched with 1% antibiotics (penicillin/streptomycin) and 10% fetal bovine serum (FBS). The cells were cultured at 37 °C in a humidified incubator with 5% CO2. For HaCaT cells, 5×10^3 cells per well were seeded into 96-well plates one day prior to treatment with increasing peptide concentrations (3.12–50 µmol L^{-1}). THP-1 monocytes were differentiated into macrophages by treating them with 100 nmol L^{-1} phorbol 12-myristate 13-acetate (PMA) for three days before administering the peptides. After a 24-hour incubation with each peptide, an MTT assay was performed. Briefly, MTT reagent (0.5 mg mL^{-1}) in phenol red-free medium (DMEM for HaCaT cells and RPMI for PMA-differentiated THP-1 cells) was added to each well (100 µL per well). The plates were incubated at 37 °C for 4 hours to facilitate the formation of insoluble formazan salts. These salts were subsequently dissolved in a solution of 0.04 mol L^{-1} hydrochloric acid (HCl) in anhydrous isopropanol, and the absorbance at 570 nm was recorded using a spectrophotometer.

Hemolysis assays

To assess the release of hemoglobin from human red blood cells (RBCs) following treatment with each peptide, human erythrocytes (blood type A-, obtained from ZenBio, using heparin-anticoagulated blood) were washed three times with PBS (pH 7.4) by centrifugation at 800 g for 10 minutes. A 1:200 dilution of RBCs (75 µL) was combined with different concentrations of peptide solution (0.78–100 µmol L^{-1}; 75 µL), and the resulting mixture was incubated at room temperature for 4 hours. After incubation, the mixture was centrifuged at 1,300 g for 10 minutes to pellet the cells and debris. The

supernatant (100 µL) from each well was moved to a new 96-well plate for absorbance analysis at 405 nm using an automated plate reader. The percentage of hemolysis was calculated by comparing the absorbance to the negative control (PBS) and the positive control (1% SDS in PBS, v/v).

$$Hemolysis\ (\%) = \frac{100 \times (Absorbance_{405\ nm\ peptide} - Absorbance_{405\ nm\ negative\ control})}{(Absorbance_{405\ nm\ positive\ control} - Absorbance_{405\ nm\ negative\ control})}$$

Circular dichroism experiments

Circular dichroism (CD) measurements were performed using a J1500 CD spectropolarimeter (Jasco) at the Biological Chemistry Resource Center (BCRC) of the University of Pennsylvania. The experiments were carried out at 25 °C, and the spectra shown are the result of averaging three accumulations. A quartz cuvette with a 1.0 mm optical path length was used, covering a wavelength range of 260–190 nm at a scanning speed of 50 nm min^{-1} and a bandwidth of 0.5 nm. Peptide samples were tested at a concentration of 50 µmol L^{-1} in various environments: water, a 60% trifluoroethanol (TFE) in water mixture, a 50% methanol (MeOH) in water mixture, and water containing sodium dodecyl sulfate (SDS) at 10 mmol L^{-1}. Baseline readings were taken before data acquisition, and a Fourier transform filter was used to minimize background noise. The secondary structure fraction values were calculated using the single-spectrum analysis tool on the BeStSel server [21]. Ternary plots were created in https://www.ternaryplot.com/ and subsequently edited.

Intracellular antimicrobial activity

HaCaT and PMA-differentiated THP-1 cells were seeded at a density of 50,000 cells per well (100 µL) in antibiotic-free medium. The cells were incubated for 2 hours with *S. aureus* ATCC 12600 at initial bacterial densities of 0.002 OD mL^{-1} for keratinocytes and 0.001 OD mL^{-1} for macrophages to allow infection. After the incubation period, gentamicin was added at a final concentration of 500 µg mL^{-1} to remove extracellular bacteria. The samples were then washed once with PBS and treated with peptides at concentrations of 12.5, 25, and 50 µmol mL^{-1}. At designated time points (0, 1, 3, and 20 hours post-treatment), the supernatants were collected, serially diluted 10-fold, and plated on LB agar for colony counting. The cellular pellets were washed with PBS, lysed using 1% Triton X-100, and similarly plated on LB agar for bacterial enumeration.

Fluorescence microscopy analysis of Cy5-Turingcin internalization in eukaryotic Cells

HaCaT and PMA-differentiated THP-1 cells were seeded on glass coverslips in 24-well plates at densities of 50,000 and 400,000 cells per well, respectively. Experimental conditions were designed to evaluate Cy5-Turingcin internalization in both infected and uninfected cells. For the infection condition, cells were incubated for 2 hours with *S. aureus* ATCC 12600 at an initial bacterial density of 0.02 OD mL^{-1} to enhance bacterial visualization under the microscope. Following bacterial exposure, cells were treated with gentamicin for an additional 2 hours as previously described. After treatment, cells were

washed with PBS and incubated with Cy5-Turingcin at concentrations of 25 and 50 µmol L^{-1}. In parallel, peptide treatments were performed on uninfected cells that were not exposed to *S. aureus*. At three time points (0, 1, and 3 hours post-treatment), cells were washed with PBS and fixed with 4% paraformaldehyde in PBS for 15 minutes at room temperature. Nuclei, cell walls, and bacteria were stained using Hoechst 33342 (1 µg mL^{-1}) and wheat germ agglutinin conjugated to Alexa Fluor 488 (WGA 488, 10 µg mL^{-1}) for 20 minutes at room temperature. The stained cells were washed twice with PBS, mounted on glass slides, and imaged using a Nikon fluorescence microscope.

Eukaryotic membrane depolarization assays using $DiBAC_4$-3

HaCaT and PMA-differentiated THP-1 cells were seeded at a density of 50,000 cells per well in 96-well plates (Corning 3904) using phenol red-free medium to minimize background fluorescence. Cellular monolayers were washed twice with 100 µL of a physiological buffer composed of 140 mmol L^{-1} NaCl, 5.4 mmol L^{-1} KCl, 1 mmol L^{-1} $CaCl_2$, 1 mmol L^{-1} $MgCl_2$, 10 mmol L^{-1} glucose, and 10 mmolL^{-1} HEPES (pH 7.4) to remove residual medium. The cells were subsequently incubated with 5 µmol L^{-1} $DiBAC_4$-3, a voltage-sensitive anionic dye (Thermo Fisher, catalog no. B438), prepared in the same buffer. This incubation was conducted at 37°C for 40 minutes to ensure adequate dye uptake and equilibration. Following incubation, the monolayers were gently washed to remove excess dye before being exposed to peptides at their CC_{50} concentrations. Fluorescence changes (λ_{ex} = 490 nm, λ_{em} = 516 nm) were recorded in real-time for 45 minutes using a microplate reader, providing a quantitative assessment of membrane depolarization dynamics.

Resistance to proteolytic degradation assay

The stability of Turingcin against enzymatic degradation was assessed by incubation in a solution containing 25% human serum in water. Specifically, the peptide, at a concentration of 7 mg mL^{-1}, was incubated with an aqueous solution of 25% human serum (Zen-Bio; healthy donor, blood type A-) for 6 hours. Samples were collected at 0.5, 1, 2, 4, and 6 hours, and each aliquot was immediately treated with 10 µL of trifluoroacetic acid for 10 minutes to halt enzymatic activity. Peptide analysis was performed using a Waters Acquity ultra-high-performance liquid chromatography-mass spectrometry (UHPLC-MS) system featuring a photodiode array detector (190–400 nm data collection) and a Waters single quadrupole detector 2 (SQD2). The setup included a Waters XBridge C_{18} column (3.5 µm, 4.6 mm x 50 mm), with a mobile phase consisting of 100% water containing 0.1% (v/v) formic acid (solvent A) and 100% acetonitrile (solvent B). Both solvents were of Fisher Optima grade. A 20 µL injection volume was used, and ionization was carried out in both positive and negative ESI modes, scanning a mass range of m/z 100–3,000. The gradient used consisted in 5-95% solvent B in 6 min. The proportion of intact peptide remaining at each time point was calculated by integrating the area under the curve of the peptide peak at the initial time (t = 0) as a reference. The experiments were done in three replicates.

Skin abscess infection mouse model

Six-week-old female CD-1 mice were anesthetized, and their backs were shaved. The area was then wiped with ethanol to eliminate any endogenous S*taphylococci* prior to creating a surface abrasion using a scalpel. An aliquot of *S. aureus* ATCC 12600 (10^7 CFU mL^{-1}; 20 μL), grown to an optical density of 0.45 at 600 nm in LB medium and washed twice with sterile PBS (pH 7.4) by centrifugation (5,000 g for 5 minutes), was applied to the scratched region. Peptide or gentamicin, each diluted to the $10\times MIC_{100}$, was administered to the wound site 2 hours post-infection. One day after the infection, the animals were euthanized, and the infected skin was excised. The tissues were weighed and processed into a homogenized mixture using a bead beater at 25 Hz for 20 minutes. The samples were then serially diluted 10-fold and plated on Mannitol Salt Agar (MSA) plates for colony-forming unit (CFU) quantification. Ten mice per group were included in the study, with two animals designated for histological analysis. All procedures were approved by the University Laboratory Animal Resources (ULAR) at the University of Pennsylvania (Protocol 806763).

Histological analysis

Skin tissues from mice, both untreated and treated with Cy5-Turingcin, were collected, fixed in 4% paraformaldehyde in PBS overnight at 4°C, and stored in 70% ethanol prior to sectioning. Tissue sections were deparaffinized and rehydrated through a series of 5-minute incubations in xylene followed by a graded ethanol series (100% to 70%).

The slides were washed in PBS, and antigen retrieval was performed in citrate buffer (10 mmol L^{-1} citrate, pH 6) at 90°C for 1 hour. After antigen retrieval, the slides were washed again in PBS and blocked with 2% BSA and 0.05% Tween 20 in PBS for 30 minutes. Sections were then incubated with WGA 488 (10 μg mL^{-1}) in 2% BSA and 0.05% Tween 20 in PBS for 30 minutes. Following this, the sections were washed twice in PBS for 3 minutes each and stained with Hoechst 33342 (1 μg mL^{-1}) in PBS for 5 minutes. The sections were then washed twice in PBS for 3 minutes each and mounted using a mounting medium. Images were acquired using Zeiss LSM 980 confocal maintained by the CDB Microscopy Core.

Peritonitis/sepsis mouse model

Six-week-old female CD-1 mice (stock number: 18679700-022) were infected intraperitoneally (I.P.) with *S. aureus* ATCC 12600 (5×10^{10} CFU mL^{-1}; 100 μL). The bacteria were cultured to an optical density of 0.5 at 600 nm in LB medium, washed twice with sterile PBS (pH 7.4) by centrifugation (3,500 g for 10 minutes), and resuspended in PBS. Mice were treated with three intraperitoneal doses of either gentamicin (460 μmol L^{-1}) or peptide (975 μmol L^{-1}), each administered in a 100 μL volume at 1.5-hour intervals, beginning 2 hours post-infection. Twenty-four hours after the final treatment, the mice were euthanized, and peritoneal fluids were collected by washing the peritoneal cavity with 5 mL of ice-cold PBS. The bacterial load in the fluids was assessed by dividing the sample into two equal fractions. For extracellular bacteria quantification, one fraction was centrifuged at 300 g for 10 minutes at 4 °C, and the supernatant was plated on MSA plates. For intracellular bacteria quantification, the second fraction was centrifuged at 1,000 rpm for 5 minutes at 4 °C, and the resulting pellet was treated with

gentamicin (500 μg mL^{-1}) for 3 hours at 37 °C to eliminate extracellular bacteria. The pellet was then washed twice with PBS, serially diluted, and plated on MSA plates. Liver, kidneys, spleen, lungs, and heart were harvested, weighed, and homogenized using a bead beater (25 Hz for 20 minutes). The homogenates were plated on MSA to evaluate bacterial dissemination to these organs. Nine mice per group were included in the study. All procedures were approved by the University Laboratory Animal Resources (ULAR) at the University of Pennsylvania (Protocol 807530).

Quantification and statistical analysis

Reproducibility of the experimental assays

Unless specified otherwise, all assays were performed in three independent biological replicates, as outlined in the figure legends and the Experimental Models and Methods sections. Cytotoxic and hemolytic activity values were calculated using non-linear regression analysis of peptide screenings across a concentration gradient, representing the concentrations required to kill 50% of the cells in the experiment. For the skin abscess and peritonitis/sepsis infection mouse models, 10 mice per group, including 2 used for histological analysis, and 9 mice per group, respectively, were employed, in accordance with protocols approved by the University Laboratory of Animal Resources (ULAR) at the University of Pennsylvania.

Statistical tests

In the intracellular antimicrobial activity assays and mouse experiments, all raw data were transformed using a log10 scale. Statistical significance at each time point was assessed independently using one-way ANOVA followed by Dunnett's test. P-values are reported for each group, with all groups compared to the untreated control group.

Statistical analysis

All calculation and statistical analyses of the experimental data were conducted using GraphPad Prism v.10.3. Statistical significance between different groups was calculated using the tests indicated in each legend.

Acknowledgments

Cesar de la Fuente-Nunez holds a Presidential Professorship at the University of Pennsylvania and acknowledges funding from the Procter & Gamble Company, United Therapeutics, a BBRF Young Investigator Grant, the Nemirovsky Prize, Penn Health-Tech Accelerator Award, Defense Threat Reduction Agency grants HDTRA11810041 and HDTRA1-23-1-0001, and the Dean's Innovation Fund from the Perelman School of Medicine at the University of Pennsylvania. Research reported in this publication was supported by the Langer Prize (AIChE Foundation), the NIH R35GM138201, and DTRA HDTRA1-21-1-0014.We thank Dr. Mark Goulian for kindly donating the following strains: *Escherichia coli* AIC221 (*Escherichia coli* MG1655 phnE_2::FRT [control strain for AIC222]) and *Escherichia coli* AIC222 (*Escherichia coli* MG1655 pmrA53

phnE_2::FRT [polymyxin resistant]), and the H-MARC core in the Center for Molecular Studies in Digestive and Liver Diseases (P30 DK050306) for kindly donating strains *Acinetobacter baumannii* ATCC 19606 and *Klebsiella pneumoniae* ATCC 13883. We would like to acknowledge the Center for Molecular Studies in Digestive and Liver Diseases (P30DK050306) and the Molecular Pathology and Imaging Core (RRID: SCR_022420) for their assistance with histology sample preparation. Additionally, we thank the Cell & Developmental Biology Microscopy Core at the Perelman School of Medicine (RRID: SCR_022373) for enabling the use of the Zeiss LSM 980 confocal microscope. We thank Gui-Shuang Ying from the Center for Preventive Ophthalmology and Biostatistics (CPOB) Consulting Service for advice on our statistical analyses. We thank the de la Fuente lab members for insightful discussions, especially Sufen Li, for her assistance with animal experiments. Figures created with Biorender.com are attributed as such. Molecules were rendered using the PyMOL Molecular Graphics System, Version 3.1.1 Schrödinger, LLC.

Funding:

National Institute of General Medical Sciences of the National Institutes of Health grant R35GM138201 (CFN)

Defense Threat Reduction Agency grant HDTRA1-21-1-0014 (CFN)

Author contributions:

Conceptualization: AC, CFN
Methodology: AC, FW, MDTT, CFN
Experimental investigation: AC, MDTT
Computational investigation: FW
Visualization: AC, FW, MDTT
Funding acquisition: CFN
Supervision: CFN
Formal analysis: AC, FW, MDTT
Writing – original draft: AC, FW, MDTT, CFN
Writing – review & editing: AC, FW, MDTT, CFN

Competing interests:

CFN provides consulting services to Invaio Sciences and is a member of the Scientific Advisory Boards of Nowture S.L., Peptidus, European Biotech Venture Builder and Phare Bio. CFN is also a member of the Advisory Board for the Peptide Drug Hunting Consortium (PDHC). The de la Fuente Lab has received research funding or in-kind donations from United Therapeutics, Strata Manufacturing PJSC, and Procter & Gamble, none of which were used in support of this work. An invention disclosure associated with this work has been filed. All other authors declare no competing interests.

Data and material availability:

Pretrained models for APEX$_{DUO}$ is available on Github (https://gitlab.com/machine-biology-group-public/apex_duo). All experimental validation data can be found in the main text or supplementary materials.

Supplementary Materials

Figs. S1 to S7
Tables S1 to S3
References

References

1. Health department Healthcare-associated Infections and Antimicrobial Resistance Programs. *Centers for Disease Control and Prevention* https://www.cdc.gov/healthcare-associated-infections/programs/index.html (2024).

2. Naghavi, M. *et al.* Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *The Lancet* 404, 1199–1226 (2024).

3. Thakur, A., *et al*. Intracellular pathogens: Host immunity and microbial persistence strategies. *Journal of Immunology Research* vol. 2019 Preprint at https://doi.org/10.1155/2019/1356540 (2019).

4. Eisenreich, W., *et al*. How viral and intracellular bacterial pathogens reprogram the metabolism of host cells to allow their intracellular replication. *Frontiers in Cellular and Infection Microbiology* vol. 9 Preprint at https://doi.org/10.3389/fcimb.2019.00042 (2019).

5. de la Fuente-Nunez, C., *et al*. Antibiotic failure: Beyond antimicrobial resistance. *Drug Resistance Updates* 71, 101012 (2023).

6. Peyrusson, F. *et al.* Intracellular Staphylococcus aureus persisters upon antibiotic exposure. *Nat Commun* 11, 2200 (2020).

7. Cesaro, A., *et al*. Deep learning tools to accelerate antibiotic discovery. *Expert Opin Drug Discov* 18, 1245–1257 (2023).

8. Gomarasca, M. *et al.* Bacterium-Derived Cell-Penetrating Peptides Deliver Gentamicin To Kill Intracellular Pathogens. *Antimicrob Agents Chemother* 61, e02545-16 (2017).

9. Derakhshankhah, H. & Jafari, S. Cell penetrating peptides: A concise review with emphasis on biomedical applications. *Biomedicine & Pharmacotherapy* 108, 1090–1096 (2018).

10. Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* 24, 1551–7 (2006).

11. Cesaro, A. *et al.* Synthetic Antibiotic Derived from Sequences Encrypted in a Protein from Human Plasma. *ACS Nano* 16, 1880–1895 (2022).

12. Bravo-Santano, N. *et al.* Intracellular Staphylococcus aureus Modulates Host Central Carbon Metabolism To Activate Autophagy. *mSphere* 3, e00374-18 (2018).

13. Benfield, A. H. & Henriques, S. T. Mode-of-Action of Antimicrobial Peptides: Membrane Disruption vs. Intracellular Mechanisms. *Front Med Technol* 2, 610997 (2020).

14. Mohd Yasin, Z. N., *et al*. Macrophage polarization in THP-1 cell line and primary monocytes: A systematic review. *Differentiation* 128, 67–82 (2022).

15. Cesaro, A., *et al*. Methods for the design and characterization of peptide antibiotics. in *Methods in Enzymology* 303–326 (Academic Press, 2022). doi:10.1016/bs.mie.2021.11.003.

16. Ma, X. *et al.* A Review of Antimicrobial Peptides: Structure, Mechanism of Action, and Molecular Optimization Strategies. *Fermentation* 10, 540 (2024).

17. Eiríksdóttir, E., *et al*. Secondary structure of cell-penetrating peptides controls membrane interaction and insertion. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1798, 1119–1128 (2010).

18. Luo, P. & Baldwin, R. L. Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 36, 8413–21 (1997).

19. Pastor, M. T., *et al*. Combinatorial approaches: A new tool to search for highly structured β-hairpin peptides. *Proceedings of the National Academy of Sciences* 99, 614–619 (2002).

20. Tulumello, D. V. & Deber, C. M. SDS Micelles as a Membrane-Mimetic Environment for Transmembrane Segments. *Biochemistry* 48, 12096–12103 (2009).

21. Micsonai, A. *et al.* BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic Acids Res* 50, W90–W98 (2022).

22. Torres, M. D. T., *et al*. Peptide Design Principles for Antimicrobial Applications. *J Mol Biol* 431, 3547–3567 (2019).

23. Ju, X. *et al.* Alanine-Scanning Mutational Analysis of Durancin GL Reveals Residues Important for Its Antimicrobial Activity. *J Agric Food Chem* 63, 6402–6409 (2015).

24. Trofimenko, E. *et al.* Genetic, cellular, and structural characterization of the membrane potential-dependent cell-penetrating peptide translocation pore. *Elife* 10, e69832 (2021).

25. de Souza-Guerreiro, T. C. *et al.* Membrane potential dynamics unveil the promise of bioelectrical antimicrobial susceptibility Testing (BeAST) for anti-fungal screening. *mBio* 15, e0130224 (2024).

26. Torres, M. D. T. *et al.* Mining for encrypted peptide antibiotics in the human proteome. *Nat Biomed Eng* 6, 1451–1451 (2022).

27. Torres, M. D. T. *et al.* Mining human microbiomes reveals an untapped source of peptide antibiotics. *Cell* 187, 5453-5467.e15 (2024).

28. Torres, M. D. T., *et al*. Peptides from non-immune proteins target infections through antimicrobial and immunomodulatory properties. *Trends Biotechnol* 0167–7799 (2024) doi:10.1016/j.tibtech.2024.09.008.

29. Wang, X. *et al.* Increased intracellular activity of MP1102 and NZ2114 against Staphylococcus aureus in vitro and in vivo. *Sci Rep* 8, 4204 (2018).

30. Sandberg, A., *et al*. Intracellular Activity of Antibiotics against *Staphylococcus aureus* in a Mouse Peritonitis Model. *Antimicrob Agents Chemother* 53, 1874–1883 (2009).

31. Cruz, G. S., *et al*. Cell-Penetrating Antimicrobial Peptides with Anti-Infective Activity against Intracellular Pathogens. *Antibiotics* 11, 1772 (2022).

32. Suzek, B. E., *et al*. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288 (2007).

33. Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* 44, D1087–D1093 (2016).

34. Pirtskhalava, M. *et al.* DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res* 49, D288–D297 (2021).

35. Shi, G. *et al.* DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic Acids Res* 50, D488–D496 (2022).

36. Wan, F., *et al*. Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nat Biomed Eng* 8, 854–871 (2024).

37. CPPsite 2.0. http://crdd.osdd.net/raghava/cppsite/.

38. Agrawal, P. *et al.* CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res* 44, D1098–D1103 (2016).

39. Wei, L., *et al*. SkipCPP-Pred: an improved and promising sequence-based predictor for predicting cell-penetrating peptides. *BMC Genomics* 18, 742 (2017).

40. Manavalan, B., *et al*. Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy. *J Proteome Res* 17, 2715–2726 (2018).

41. Singh, S. *et al.* SATPdb: a database of structurally annotated therapeutic peptides. *Nucleic Acids Res* 44, D1119–D1126 (2016).

42. Shengjia Zhao, *et al*. InfoVAE: Information Maximizing Variational Autoencoders. (2018) doi:https://doi.org/10.48550/arXiv.1706.02262.

43. Pirtskhalava, M. *et al*. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res* 49, D288–D297 (2021).

44. Boukamp, P. *et al*. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J Cell Biol* 106, 761–771 (1988).

Figures

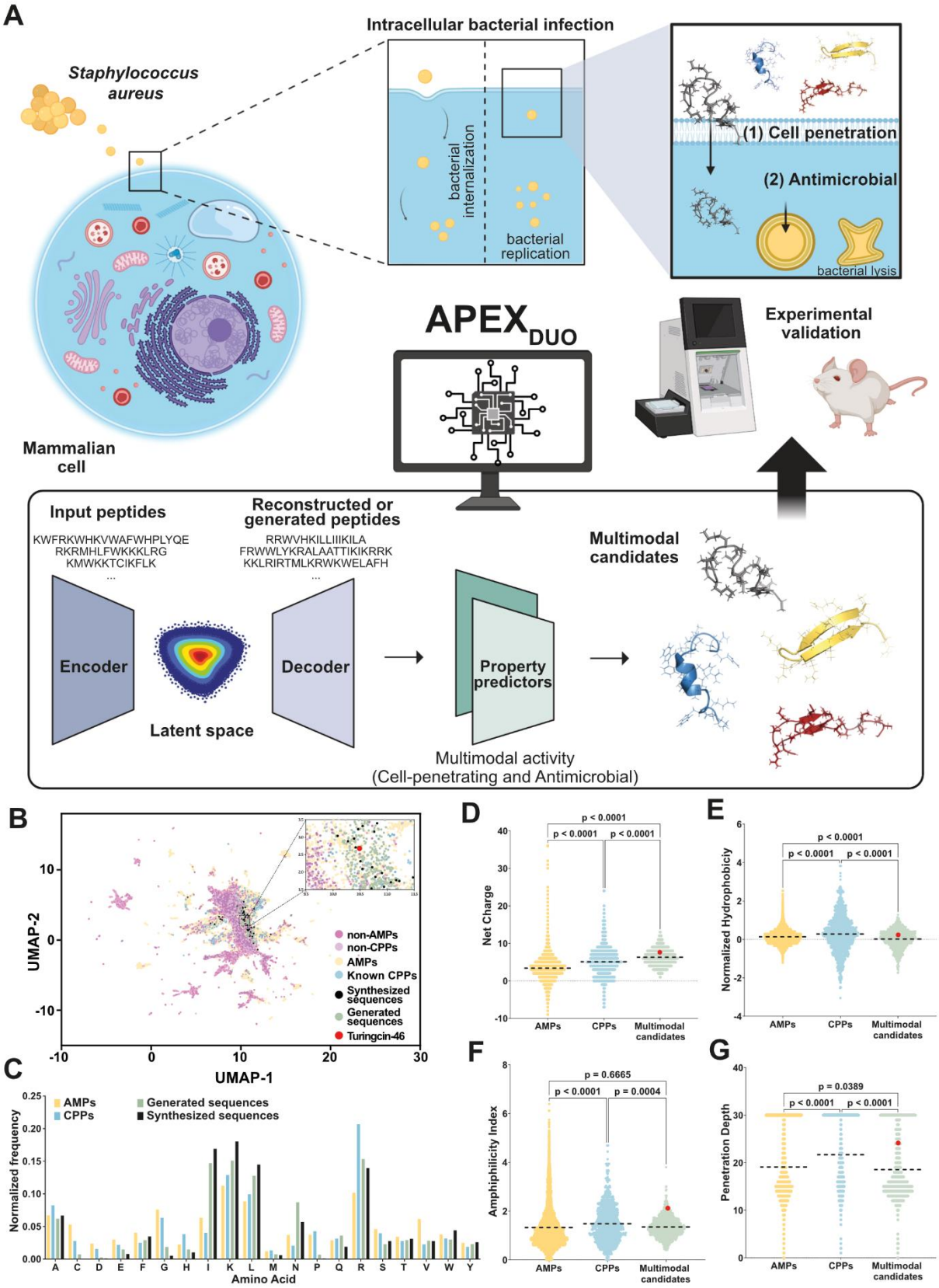

Fig. 1. Multimodal peptide generation using deep learning. (A) A deep learning workflow capable of generating peptide candidates that can penetrate eukaryotic cells and

eliminate intracellular *S. aureus*. A variational autoencoder (VAE) was used to generate novel peptide sequences. The resulting candidates were then filtered using predictive deep learning models to select those with high antimicrobial and cell-penetrating potentials for further experimental validation. The peptide structures shown were rendered using the PyMOL Molecular Graphics System (version 3.0, Schrödinger, LLC). Figure created with BioRender.com. **(B)** Distribution of different peptide classes in ESM2-derived sequence space. Sequences from publicly available AMPs and CPPs, as well as multimodal candidates generated by APEXDUO, were embedded into a 2D space after dimensionality reduction via a sequence similarity matrix. This analysis reveals that the newly generated multimodal candidates occupy a region distinct from non-AMP/CPP sequences, yet overlap with the known AMP/CPP space, suggesting the potential for dual functionality. **(C)** Comparison of amino acid frequencies in Turingcins (both generated and synthesized) against publicly available AMPs and CPPs. The multimodal peptides show a unique compositional profile, including a relative enrichment of lysine, isoleucine, leucine, and asparagine residues. **(D–G)** Comparative analysis of physicochemical properties among the multimodal candidates, AMPs, and CPPs. Properties include **(D)** net charge, **(E)** normalized hydrophobicity, **(F)** amphiphilicity index, and **(G)** penetration depth. Normalized hydrophobicity influences peptide-membrane interactions, net charge modulates electrostatic attractions with negatively charged bacterial membranes, amphiphilicity index correlates with the peptide's mechanism of membrane engagement, and penetration depth reflects how deeply the peptide integrates into lipid bilayers. Turingcin-46, the most promising multimodal candidate identified in this study, is indicated with a red dot. For panels **D–G**, statistical significance was determined via two-tailed t-tests followed by Mann–Whitney tests; p values are shown on the graphs. The dashed line within each distribution denotes the mean for that group.

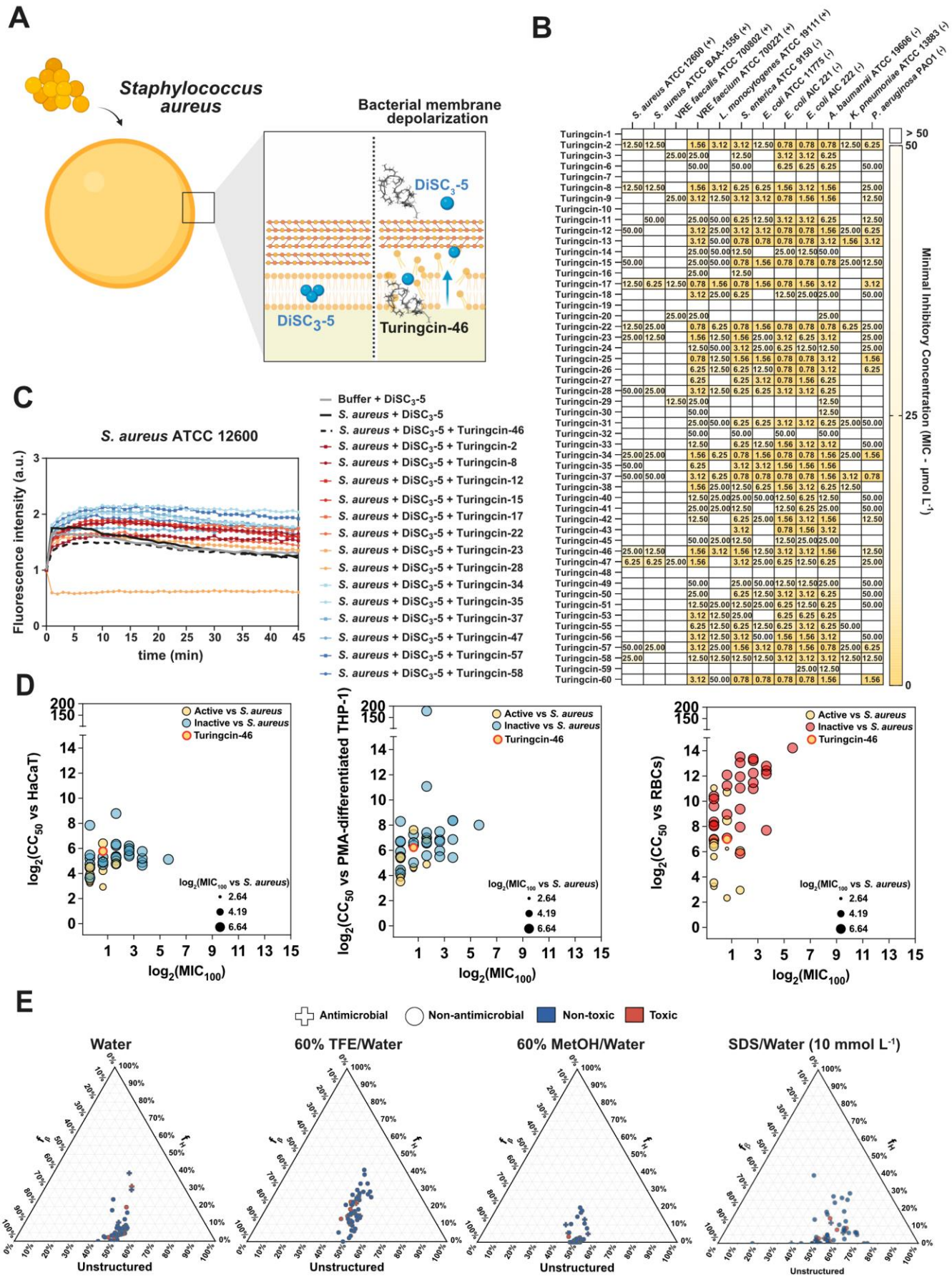
A
Staphylococcus aureus
Bacterial membrane depolarization
DiSC3-5
DiSC3-5
Turingcin-46
B
S. aureus ATCC 12600 (+)
S. aureus ATCC BAA-1556 (+)
VRE faecalis ATCC 700802 (+)
VRE faecium ATCC 700221 (+)
L. monocytogenes ATCC 19111 (+)
S. enterica ATCC 9150 (-)
E. coli ATCC 11775 (-)
E. coli AIC 221 (-)
E. coli AIC 222 (-)
A. baumannii ATCC 19606 (-)
K. pneumoniae ATCC 13883 (-)
P. aeruginosa PAO1 (-)
Minimal Inhibitory Concentration (MIC - µmol L-1)
> 50
50
25
0
C
S. aureus ATCC 12600
Fluorescence intensity (a.u.)
time (min)
Buffer + DiSC3-5
S. aureus + DiSC3-5
S. aureus + DiSC3-5 + Turingcin-46
S. aureus + DiSC3-5 + Turingcin-2
S. aureus + DiSC3-5 + Turingcin-8
S. aureus + DiSC3-5 + Turingcin-12
S. aureus + DiSC3-5 + Turingcin-15
S. aureus + DiSC3-5 + Turingcin-17
S. aureus + DiSC3-5 + Turingcin-22
S. aureus + DiSC3-5 + Turingcin-23
S. aureus + DiSC3-5 + Turingcin-28
S. aureus + DiSC3-5 + Turingcin-34
S. aureus + DiSC3-5 + Turingcin-35
S. aureus + DiSC3-5 + Turingcin-37
S. aureus + DiSC3-5 + Turingcin-47
S. aureus + DiSC3-5 + Turingcin-57
S. aureus + DiSC3-5 + Turingcin-58
D
Active vs S. aureus
Inactive vs S. aureus
Turingcin-46
log2(CC50 vs HaCaT)
log2(CC50 vs PMA-differentiated THP-1)
log2(CC50 vs RBCs)
log2(MIC100)
log2(MIC100 vs S. aureus)
2.64
4.19
6.64
E
Antimicrobial
Non-antimicrobial
Non-toxic
Toxic
Water
60% TFE/Water
60% MetOH/Water
SDS/Water (10 mmol L-1)
Unstructured

Fig. 2. Antimicrobial activity, cytotoxicity, and secondary structure profiles of Turingcins. (A) Schematic overview of how Turingcins exert antimicrobial activity against *S. aureus*, including the induction of membrane depolarization. **(B)** Heat map showing the minimal inhibitory concentration (MIC_{100}) values of active Turingcins generated by $APEX_{DUO}$ against clinically significant Gram-positive (+) and Gram-negative (-) bacterial strains, including antibiotic-resistant isolates. Cells were treated with serial dilutions of each peptide (0.78–50 μmol L^{-1}) at 37°C, and bacterial growth was measured after 24 hours via optical density at 600 nm. MIC_{100} values indicate the lowest concentration that completely inhibited growth, reported as the mode of two independent replicates. **(C)** Relative fluorescence changes of *S. aureus* membranes treated with Turingcins in the presence of the voltage-sensitive dye $DiSC_3$-5, indicating peptide-induced membrane depolarization. All experiments were performed in three independent replicates (see also **Fig. S2a**) **(D)** Scatterplots correlating antimicrobial activity (MIC_{100}) with cytotoxicity [CC_{50} for human keratinocytes (HaCaT), PMA-differentiated human monocytes (THP-1), and HC_{50} for human red blood cells (RBCs)]. CC_{50} and HC_{50} values were derived from dose-response curves obtained via non-linear regression analysis (see also **Figs. S2B**, **C**, and **D**). **(E)** Ternary plots depicting the secondary structure composition (α-helix, β-sheet, and disordered) of each peptide at 50 μmol L^{-1} in different solvents: water, 60% trifluoroethanol (TFE) in water, 60% methanol (MeOH) in water, and sodium dodecyl sulfate (SDS, 10 mmol L^{-1}) in water. Secondary structure fractions were estimated using the BeStSel server [21] (see also **Fig. S3**). Figure created with BioRender.com.

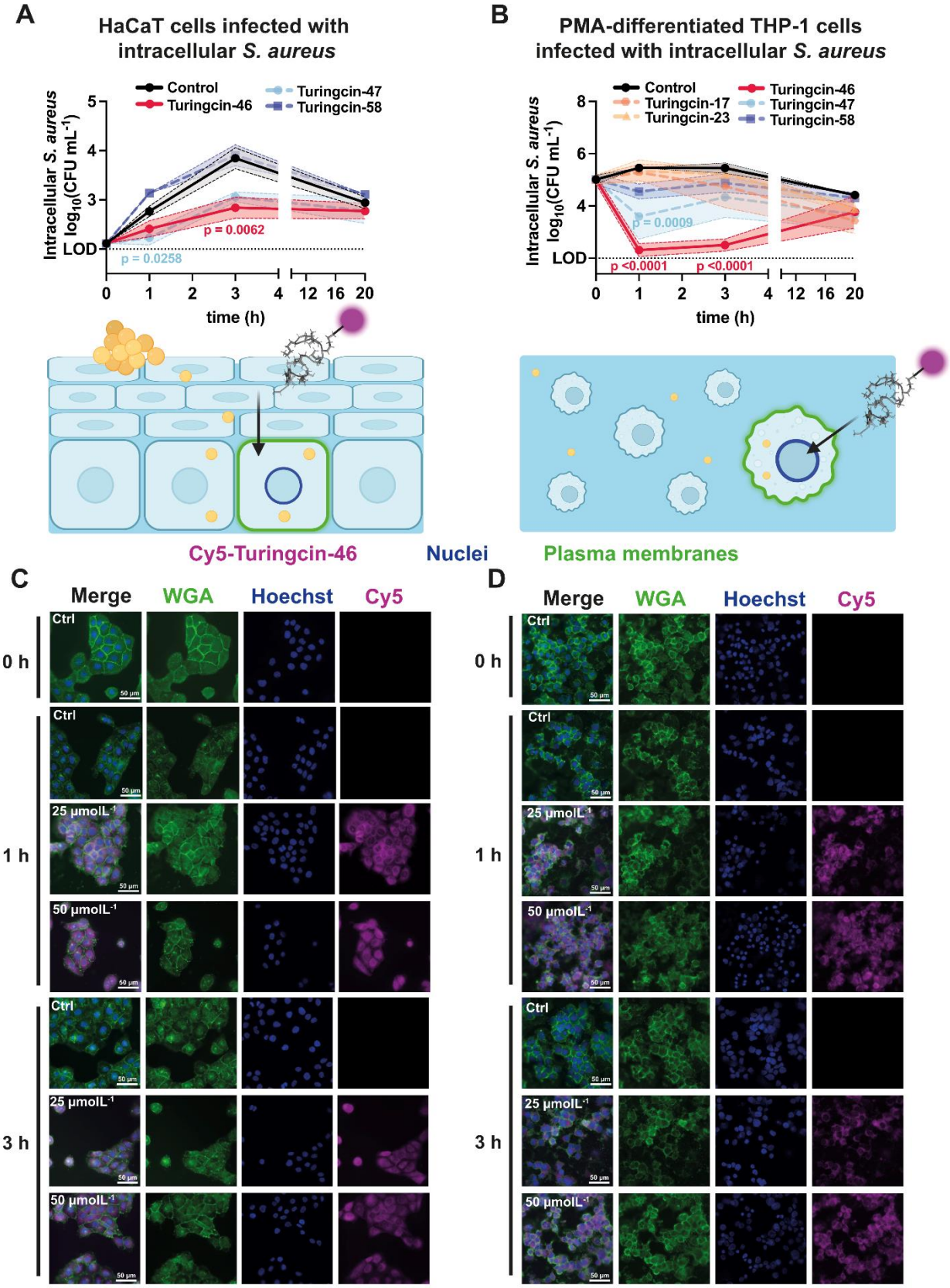

Fig. 3. Intracellular antimicrobial activity of Turingcins. (A, B) Time-course analysis of the intracellular antimicrobial activity of Turingcin peptides at non-cytotoxic concentrations against *S. aureus* ATCC 12600 in HaCaT keratinocytes and PMA-

differentiated THP-1 macrophages. Intracellular bacterial loads were quantified at 0, 1, 3, and 20 hours post-treatment. A minimum of two experiments was performed for all conditions, with the lead compound (Turingcin-46) tested in three replicates. **(C, D)** Representative fluorescence microscopy images of HaCaT and PMA-differentiated THP-1 cells treated with Cy5-labeled Turingcin-46, captured at 0-, 1-, and 3-hours post-treatment. Following peptide incubation, samples were fixed with 4% PFA and stained with WGA 488 (10 $\mu g\,mL^{-1}$) to visualize plasma membranes and Hoechst 33342 (1 $\mu g\,mL^{-1}$) for nuclei visualization. Each experiment was performed in biological duplicate, with two technical replicates each. Figures were created using BioRender.com.

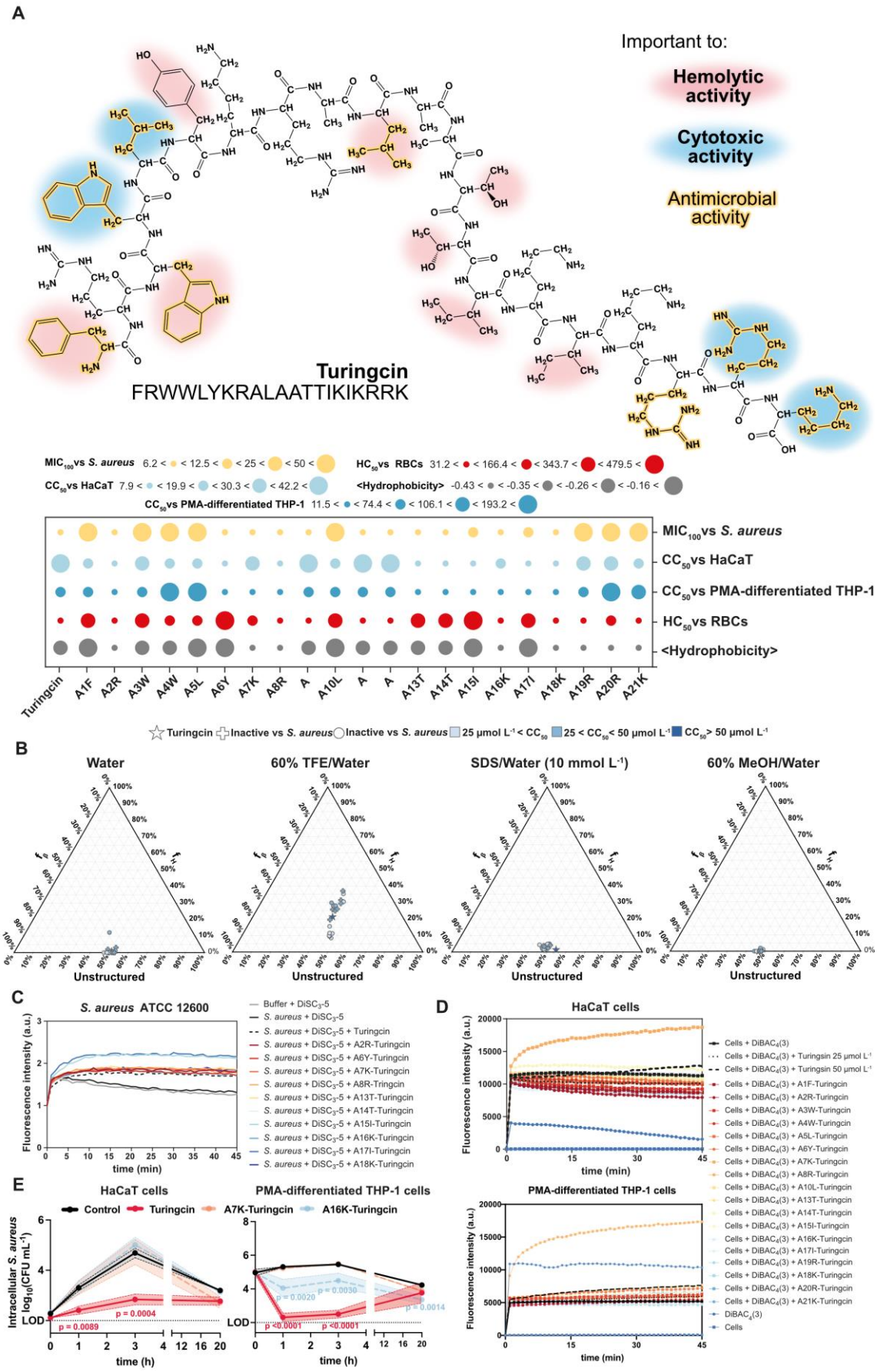
A
Important to:
Hemolytic activity
Cytotoxic activity
Antimicrobial activity
Turingcin
FRWWLYKRALAATTIKIKRRK
MIC100 vs S. aureus
CC50 vs HaCaT
CC50 vs PMA-differentiated THP-1
HC50 vs RBCs
<Hydrophobicity>
Turingcin
A1F
A2R
A3W
A4W
A5L
A6Y
A7K
A8R
A
A10L
A
A
A13T
A14T
A15I
A16K
A17I
A18K
A19R
A20R
A21K
B
Water
60% TFE/Water
SDS/Water (10 mmol L-1)
60% MeOH/Water
Unstructured
C
S. aureus ATCC 12600
Fluorescence intensity (a.u.)
time (min)
D
HaCaT cells
PMA-differentiated THP-1 cells
E
HaCaT cells
PMA-differentiated THP-1 cells
Control
Turingcin
A7K-Turingcin
A16K-Turingcin
Intracellular S. aureus log10(CFU mL-1)
LOD
time (h)

Fig. 4. Structural and functional elucidation of Turingcin. (A) Chemical structure of Turingcin, with residues influencing antimicrobial activity (yellow), cytotoxicity (blue), and hemolytic effects (red) highlighted. MIC_{100} (against *S. aureus* ATCC 12600), CC_{50} (against HaCaT keratinocytes and PMA-differentiated THP-1 macrophages), HC_{50}, and normalized hydrophobicity values are presented in a bubble plot to compare the wild-type Turingcin sequence with 18 alanine-substituted variants (see also **Figs. S5A-D**). Normalized values (ranging from 0 to 1) were used for visualization, while experimental data and normalized hydrophobicity values are provided in the legend. **(B)** Ternary plots showing the secondary structure composition of each peptide (50 $\mu mol\ L^{-1}$) in different solvents: water, 60% trifluoroethanol (TFE) in water, 60% methanol (MeOH) in water, and sodium dodecyl sulfate (SDS, 10 $mmol\ L^{-1}$) in water. Secondary structures were estimated using the BeStSel server [21] (see also **Fig. S6**). **(C)** Membrane depolarization of *S. aureus* cells treated with Turingcin and alanine-substituted derivatives that retained antimicrobial activity. Changes in fluorescence intensity were measured using $DiSC_3$-5-dye at MIC_{100} concentrations (see also **Fig. S5E**). **(D)** Effects on eukaryotic cell membrane potential (depicted as depolarization or hyperpolarization) for HaCaT and PMA-differentiated THP-1 cells treated with Turingcin and alanine variants, evaluated using $DiBAC_4$-3 dye at CC_{50} concentrations (see also **Figs. S5F-G**). **(E)** Time-course analysis (0, 1, 3, and 20 hours) of intracellular antimicrobial activity of Turingcin, A7K-Turingcin, and A16K-Turingcin at non-cytotoxic concentrations against intracellular *S. aureus* in HaCaT keratinocytes and PMA-differentiated THP-1 macrophages.
All experiments were carried out in three independent replicates.

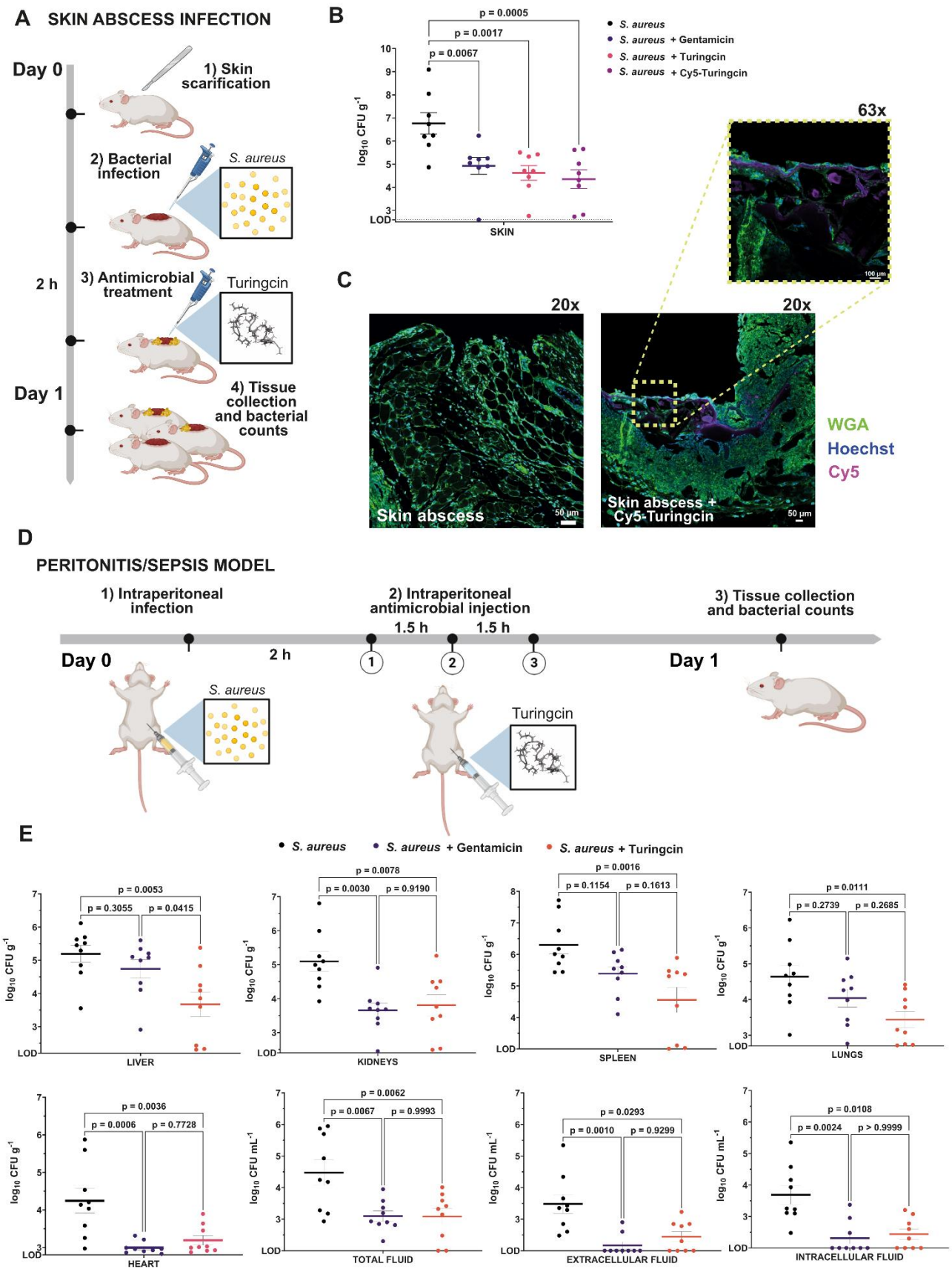

Fig. 5. Anti-infective activity of Turingcin in animal models. (A) Schematic representation of the skin abscess mouse model used to evaluate the anti-infective

efficacy of Turingcin against *S. aureus* ATCC 12600. **(B)** In the skin abscess model, a single post-infection administration of Turingcin significantly reduced bacterial load compared to the untreated control. **(C)** Representative confocal laser scanning microscopy (CLSM) images of excised skin tissues from untreated mice and those treated with Cy5-Turingcin. Deparaffinized tissue sections were stained with WGA 488 (10 $\mu g\ mL^{-1}$) to visualize plasma membranes and Hoechst 33342 (1 $\mu g\ mL^{-1}$) for nuclei. **(D)** Schematic representation of the peritonitis/sepsis mouse model, in which Turingcin was administered intraperitoneally. **(E)** Anti-infective efficacy against *S. aureus* ATCC 12600 was evaluated one day after injection of Turingcin, demonstrating significant efficacy in in multiple organs (liver, kidneys, spleen, lungs, heart) and peritoneal fluids (total, extracellular, intracellular). Compared to the untreated control, Turingcin treatment significantly decreased bacterial loads. Statistical significance in **(B, E)** was determined using one-way ANOVA followed by Dunnett's test; p values are shown in the graphs. The solid line within each distribution represents the mean. Panels **(A)** and **(D)** were created with BioRender.com.